

Accounting for observation errors in image data assimilation

By VINCENT CHABOT^{1,2,3}, MAELLE NODET^{3,2,1}, NICOLAS PAPADAKIS^{4,5} and ARTHUR VIDARD^{1,2,3*}, ¹INRIA, France; ²CNRS, LJK, F-38000 Grenoble, France; ³University Grenoble Alpes, LJK, F-38000 Grenoble, France; ⁴CNRS, IMB, UMR 5251, F-33400 Talence, France; ⁵University Bordeaux, IMB, UMR 5251, F-33400 Talence, France

(Manuscript received 20 December 2013; in final form 18 December 2014)

ABSTRACT

This paper deals with the assimilation of image-type data. Such kinds of data, such as satellite images, have good properties (dense coverage in space and time), but also one crucial problem for data assimilation: they are affected by spatially correlated errors. Classical approaches in data assimilation assume uncorrelated noise, because the proper description and numerical manipulation of non-diagonal error covariance matrices is complex. This paper proposes a simple way to provide observation error covariance matrices adapted to spatially correlated errors. This is done using various image transformations: multiscale (Wavelets, Fourier, Curvelets), gradients, and gradient orientations. These transformations are described and compared to classical approaches, such as pixel-to-pixel comparison and observation thinning. We provide simple yet effective covariance matrices for each of these transformations, which take into account the observation error correlations and improve the results. The effectiveness of the proposed approach is demonstrated on twin experiments performed on a 2-D shallow-water model.

Keywords: variational data assimilation, image assimilation, observation operator, multiscale transforms, approximate covariance matrices, correlated observation errors

1. Introduction

1.1. Motivation

In the last 30 years, data assimilation techniques have become very popular in geophysics. They aim to combine in an optimal way, through priors on the involved errors, different kinds of information in order to provide an accurate estimate of the geophysical state.

Because of the complexity to describe accurately the dynamic of the ocean or the atmosphere, mathematical models are in practice idealised and simplified representations of the reality. Observations are then necessary to monitor the evolution of these geophysical states. Observations used in operational systems historically come from synoptic data. Those data are collected by stations, aircrafts, radio-sounding, balloons, and drifters. The distribution of such observations is sparse and heterogeneous both in space and time.

Since the end of the 1970s, many satellites have been launched to improve our knowledge of the atmosphere and of the oceans. Geostationary satellites provide, among other data, photographic images of the earth system. Sequences of such images show the dynamical evolution of identified meteorological or oceanic ‘objects’: fronts, clouds, eddies, vortices and so on. The human vision can easily detect the dynamics in this kind of image sequence, and it clearly has a strong predictive potential. This aspect is favoured by the fact that image data, contrary to many other measurements, are dense in space and time. Indeed, the spatial resolution of current METEOSAT satellites is close to 1 km and they produce a full image every 15 min. This frequency will be improved up to one every 10 min (and even every 2.5 min for Europe only) for the upcoming third satellite generation. It implies a huge quantity of information which can be seen as an asset but also induces difficulties for the assimilation system that has to cope with such an amount of data.

Satellite data are currently used in an operational data assimilation system, mainly through the assimilation of the radiance measured by the satellite at each pixel of the image.

*Corresponding author.
email: Arthur.Vidard@imag.fr

They are related to physical quantities such as surface temperature, sea surface height, cloud pressure, and chlorophyll concentration. However, studies such as Desroziers et al. (2009) have shown that, although radiances measured by satellite have an important impact on the assimilation, each radiance observation has a small impact compared to that of one in situ observation. This phenomenon is due to the fact that they are integrated measures. It is thus difficult to have an accurate representation of the observation errors, and moreover these errors are highly correlated. As a consequence, the prescribed error statistics associated to radiance measurements are artificially inflated in operational systems.

In practice only a tiny percentage (about 3–5%) of total satellite (from polar orbiting and geostationary) data are used in operational numerical weather prediction (NWP) systems and given low confidence within the assimilation systems. Considering the cost of satellite observing systems (the cost of the launch of the Meteosat Third Generation is estimated at around 2.5 billion Euros) and of the infrastructures required for the collection of the data itself, improving their impact on forecasting systems is an important topic of research.

One important scientific issue related to satellite data is to study what information can, and should, be assimilated from images. In the literature, several methods dedicated to the assimilation of image sequences have been proposed. Two classes of approaches can be distinguished. The assimilation of *pseudo-observation* consists of first extracting dynamical information from images and then use the result as observation, whereas the *direct observation* relies on a single assimilation of the image data.

1.2. Pseudo-observations

The radiance measured at each pixel is relevant information, but it does not give any detail on the structures, such

as the fronts of geophysical entities, that can be observed in images. When looking at a satellite image sequence, the human eye notices the evolution of structures, through the deformation of isophote lines. Structure information is even available when looking at individual images, as illustrated by the different satellite observations of the ocean presented in Fig. 1.

In NWP, this information on the dynamic is currently assimilated through so-called atmospheric motion vectors (AMVs). Their aim is to estimate the motion of some identified structures from one image to another using correlation techniques (Lucas and Kanade, 1981). The resulting vector field is then assimilated as wind data. More complex strategies have also been designed by Michel (2011) in order to characterise specific clouds, track their motion along the image sequence and assimilate their Lagrangian trajectories.

Due to their really indirect nature and the complexity of the pre- and post-processing [a thorough description of these processes can be found in Schmetz et al. (1993) and Nieman et al. (1997)], describing the errors associated to such sparse wind data is not straightforward. In particular, they are correlated, so that complex observation error covariance matrices have to be built or the errors have to be significantly inflated, and therefore it will reduce their impact. Bormann et al. (2003) found statistically significant spatial error correlations on scales up to about 800 km, that are moreover strongly anisotropic.

Generally, AMVs are thought to be very useful. However, as shown by Cardinali (2009), such observations can in some cases have a negative impact on assimilation.

Dense motion estimation has also been considered for assimilating the image dynamics. The motion between two consecutive images can indeed be computed through image processing techniques based on dense optical flow estimation (Horn and Schunck, 1981). The resulting 2-D vector fields can therefore be associated to the velocity at sea surface or cloud altitude and assimilated as pseudo-observation, as

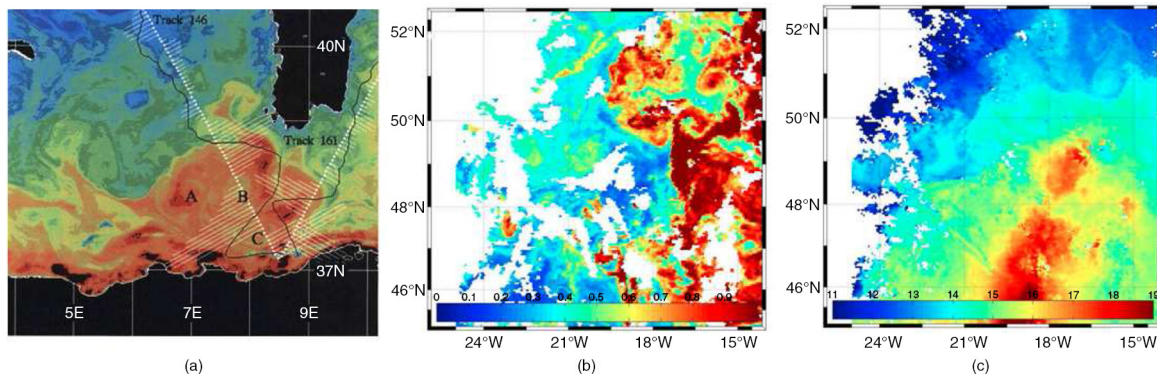


Fig. 1. Different examples of satellite images. (a) Altimetric reconstruction from JASON satellite data. (b) Ocean colour/Chlorophyll is from the MODIS captor of ENVISAT satellite. (c) Sea surface Temperature from the MODIS captor of ENVISAT satellite.

proposed by Korotaev et al. (2008) and Papadakis and Mémin (2007).

Other pseudo-observation methods based on image gradients assimilation have been proposed. Gradients contain pertinent information which has been successfully used for front tracking in oceanographic image sequences (Ba and Fablet, 2010) and assimilation of wave directions (Aouf et al., 2006).

The pseudo-observation approach nevertheless suffers from the difficulty to model the observation errors. Indeed, the error due to the image processing technique itself cannot generally be quantified accurately. For instance, the dense optical flow methods that estimate a 2-D velocity from scalar images involve an ill-posed problem and an artificial regularisation of the vector field is required. Such regularisation introduces errors in the estimation which are difficult to model. As a consequence, when dealing with pseudo-observations, it is hard to distinguish image processing errors from the ones inherent to the satellite acquisition process.

1.3. Direct comparison

A direct comparison between observations and model is also possible. For instance, the radiance observation are directly assimilated (Köpken et al., 2004). This is made possible thanks to the use of a specific observation operator named RTTOV [see, for instance, Matricardi et al. (2004)] which enable to compare directly the model variable to the observation.

In this work, we used a data assimilation method aiming at getting information on velocity field thanks to images observations. For this kind of assimilation, a direct comparison between image temporal variation and state velocity has been studied by Papadakis and Mémin (2008), through the introduction of a non-linear observation equation based on the optical flow constraint equation. This overcomes the problem of artificial regularisation by only considering the photometric information. Another approach has also been developed by Titaud et al. (2010) where observed images are compared directly to a synthetic image derived from the model state variable.

More recently, another approach using image gradients has been proposed by Titaud et al. (2011) and continued by Gaultier et al. (2013). The idea here consists of extracting a feature map from the dynamic that can be compared with a gradient map computed on images. The feature map is composed of Lyapunov exponents or vectors (D'Ovidio et al., 2009) obtained through an integration of the position of particles added to the model. More precisely, the computation consists of perturbing a particle position and measuring the direction and the amplitude of the deviation after a finite integration time. In oceanography, this information is pertinent since strong Lyapunov coefficients are corre-

lated with chlorophyll discontinuities at the surface, that can be observed from satellite images. The process nevertheless relies on a binarisation of both Lyapunov information and image gradients, which makes its integration in data assimilation systems difficult, as it involves non-differentiable processes and degrades the information.

1.4. Specification of observation errors

The specification of observation errors statistics is an important topic in data assimilation. Indeed, they define the weight of each observation and the way the analysis can differ from the observation set (through the error covariances).

Estimating the error statistic of an observation set is a mathematically difficult problem and has seldom been considered thoroughly in data assimilation. Desroziers et al. (2005) introduced a method in order to diagnose the consistency of background and observation error variances and cross-correlations under the hypothesis that the assimilation scheme relies on statistical linear estimation theory. On top of these estimation difficulties, introducing the information about cross-correlation in the assimilation scheme is a complex task due to the size of the problem and the need to invert the obtained covariance matrix. To circumvent this complexity, Stewart et al. (2013) proposed to use Markov matrices in order to model correlations. The inverse of such matrices are tridiagonal and need the storage of only one parameter. They show that, on 1-D synthetic test cases, using information on correlation (or even approximation of this information) leads to better results than considering the observation errors uncorrelated.

The specification of observation error for image-type observation is a challenge. Whatever the approach used (pseudo-observation or direct comparison) the usual assumption of uncorrelated errors between observations is clearly invalid. Additionally, the size of the data is likely to make the handling of the error covariance matrices cumbersome. The main topic of this paper is to propose and compare new approaches to specify non-trivial observation error correlations in a computationally efficient way.

1.5. Proposed approach and overview of the paper

This paper compares different observation transformations dedicated to image-type observation. Two of them have already been proposed in the literature: the classical one-to-one pixel comparison [see, for instance, Köpken et al. (2004)] and the comparison in a Curvelet space (Titaud et al., 2010) extended in this paper to the comparison in an orthonormal Wavelet and Fourier spaces. Two new transformations based on image gradients and gradient orientations are also proposed. Their relative merits are discussed

with a special attention on the description of the corresponding error statistics specification. In particular, when the observation noise is correlated in space, the usual diagonal approximation for the observation error covariance matrix is not sufficient for the pixel-to-pixel comparison, while the use of other transformations allow to continue using this approximation.

Some basic notations on variational data assimilation are first recalled in Section 2. Observation transformations are then defined in Section 3. The specification of observation error statistics for each transformation is discussed in Section 4 and non-trivial error covariance matrices for the assimilation are proposed. Finally Section 5 presents the experimental framework and some numerical results.

2. Variational data assimilation

Variational data assimilation is a technique based on optimal control theory which seeks to estimate initial or boundary conditions of a given system. This is done through the minimisation of a cost function accounting for the observations and model equations of this system.

Let \mathcal{V} be the state space identified to its dual defined over $\Omega \subset \mathbb{R}^n$ (e.g. $V = H^1(\Omega)$). The evolution of the state variable $X \in W(t_0, t_f) = \{f | f \in L^2(t_0; t_f; \mathcal{V})\}$ is assumed to be described through a (possibly non-linear) differential dynamical model $\mathcal{M} : \mathcal{V} \rightarrow \mathcal{V}$:

$$\begin{cases} \partial_t X(X_0, t) + \mathcal{M}(X(X_0, t)) = 0 \\ X(X_0, t_0) = X_0 \end{cases} \quad (1)$$

where $X_0 \in \Omega$ is a control parameter. We then consider noisy observations $Y \in \mathcal{O}$ of the state variables, where $\mathcal{O} \subset \mathbb{R}^m$ is the observation space. These observations may belong to a different space from the state variable. We will nevertheless assume that there exists a differentiable operator $\mathcal{H} : \mathcal{V} \rightarrow \mathcal{O}$, that maps the variable space to the observation space. The control problem states as follows: find the control parameter X_0 which best fits to observations over the time range $[t_0, t_f]$. To do so, we form and minimise the following cost function:

$$J^o(X_0) = \frac{1}{2} \int_{t_0}^{t_f} \|Y(t) - \mathcal{H}(X(X_0, \mathbf{x}, t))\|_{\mathbf{R}}^2 dt, \quad (2)$$

where \mathbf{R} is the observation error covariance matrix and the norm $\|\cdot\|_{\mathbf{R}}$ is defined by

$$\|u\|_{\mathbf{R}}^2 = u^T \mathbf{R}^{-1} u \quad (3)$$

for all $u \in \mathcal{O}$. It is classically known that this problem is ill-posed. Therefore, a regularisation term, involving a so-called background state, is added to the cost function:

$$J(X_0) = J^o(X_0) + J^b(X_0), \quad J^b(X_0) = \|X_0 - X^b\|_{\mathbf{B}}^2 \quad (4)$$

where \mathbf{B} is the background error covariance matrix. In this paper, it is based on the generalised diffusion equation as proposed by Weaver and Courtier (2001).

The minimum of J is sought using a gradient descent algorithm and its gradient is generally computed using an adjoint method as advocated by Le Dimet and Talagrand (1986).

3. Direct assimilation of image

A general framework needed to directly assimilate images is described in Section 3.1. Section 3.2.1 and 3.2.2 recall the pixel comparison and the multiscale transformation. Then Sections 3.2.3 and 3.2.4 introduce new transformations based on image gradients.

3.1. General framework

Let y^o be an image observation of a direct or indirect output variable of the mathematical model.

The functional to minimise reads:

$$J^o(X_0) = \sum_{t_i=t_0}^{t_f} \|\mathbf{A}(\mathcal{H}(X_{t_i})) - \mathbf{A}(y_{t_i}^o)\|_{\mathbf{R}_A}^2 \quad (5)$$

where X denotes the unknown state initial conditions that we want to estimate, including, for instance, the initial state velocity w_0 . \mathbf{A} represents a transformation ν_T that can be applied on the image and its model counterpart, and \mathbf{R}_A the associated observation error covariance matrix.

3.2. Observation transformations

This section describes transformations \mathbf{A} used to compare the images and their model equivalents.

3.2.1. Pixel. A first idea consists of defining \mathbf{A} as the identity function, so that the image comparison is performed in the pixel space. The functional to minimise is:

$$J_{pix}^o(X_0) = \sum_{t_i=t_0}^{t_f} \|\mathcal{H}(X_{t_i}) - y_{t_i}^o\|_{\mathbf{R}_{pix}}^2, \quad (6)$$

where \mathbf{R}_{pix} stands for the covariance matrix in this space.

3.2.2. Multiscale. In order to model the spatial information, an idea is to consider a multiscale representation of an image. Titaud et al. (2010) proposed to use the curvelet representation defined in Candes et al. (2006). The very same process can be done using other multiscale representations.

In this paper, we propose to use also Fourier transform and orthonormal periodic wavelet transform. This choice is due to the fact that Fourier and wavelet bases involve faster transforms and are easier to work with. The Fourier, wavelet and curvelet transforms used in this paper come from MCAlab toolbox¹ (see Fadili et al., 2010).

These transforms rely on the fact that an image y^o can be decomposed as:

$$y^o = \sum_{j,k,l} \langle y^o, \varphi_{j,k,l} \rangle \varphi_{j,k,l}$$

where $\{\varphi_{j,k,l}\}_{j,k,l}$ are the elements of the curvelet tight frame, the orthonormal wavelet or the Fourier bases. The parameters j,k,l , specify the scale, the position and the orientation of each element. These coefficients are then compared with their synthetic counterpart. Hence, the functional to minimise reads:

$$J_{\mathbf{A}_m}^o(X_0) = \sum_{t_i=t_0}^{t_f} \left\| \mathbf{A}_m(\mathcal{H}(X_{t_i})) - \mathbf{A}_m(y_{t_i}^o) \right\|_{\mathbf{R}_m}^2, \quad (7)$$

where \mathbf{A}_m stands for the multiscale transform and \mathbf{R}_m for the associated error covariance matrix. Notice that we have a Parseval equality for the multiscale transforms, so that the norm $\|\mathbf{A}_m(\mathcal{H}(X)) - \mathbf{A}_m(y^o)\|_{L^2}$ is equal to the pixel space $\|\mathcal{H}(X) - y^o\|_{L^2}$ norm.

3.2.3. Gradient. Considering that the pertinent information on the flow dynamics is localised on image singularities, we also propose to compare 2-D gradients $\nabla = [\partial_x \partial_y]^T$ of the image and its model counterpart. Hence, the observation \mathbf{A}_∇ transformation is defined as $\mathbf{A}_\nabla(y^o) = \nabla y^o$.

As a result the cost function becomes:

$$J_{\nabla}^o(X_0) = \sum_{t_i=t_0}^{t_f} \left\| \nabla \mathcal{H}(X_{t_i}) - \nabla y_{t_i}^o \right\|_{\mathbf{R}_\nabla}^2, \quad (8)$$

where \mathbf{R}_∇ is the covariance matrix relative to the observation errors \mathbf{A}_∇ .

The corresponding tangent linear transformation can be obtained for a small perturbation γ around q as:

$$\partial_q \mathbf{A}_\nabla(\gamma) = \nabla \gamma.$$

The adjoint observation transformations applied to a vector u reads:

$$(\partial_q \mathbf{A}_\nabla)^* u = -\nabla \cdot u$$

Remark 1. There exist an infinity of different images that share the same gradient map. Indeed, the family

$y_c^o = y^o + c$, where c is a global constant, is such that $\nabla y_c^o = \nabla y^o$.

3.2.4. Angular. A last transformation can be introduced to focus on structure orientations. The idea is to compare the normalised 2-D gradients $\nabla = [\partial_x \partial_y]^T$ of the image and its model counterpart, by measuring their angular difference. We then measure the misfit as

$$\frac{\nabla \mathcal{H}(X)}{\|\nabla \mathcal{H}(X)\|} - \frac{\nabla y^o}{\|\nabla y^o\|},$$

with the L^2 norm: $\|\nabla y^o\| = \sqrt{(y_x^o)^2 + (y_y^o)^2}$. In order to avoid an ill-posed problem for null vectors, we rather define the misfit as:

$$\frac{\nabla \mathcal{H}(X)}{\|\nabla \mathcal{H}(X)\|_\epsilon} - \frac{\nabla y^o}{\|\nabla y^o\|_\epsilon},$$

where $\|\nabla y^o\|_\epsilon = \sqrt{(y_x^o)^2 + (y_y^o)^2 + \epsilon^2}$ and $\epsilon > 0$ is chosen as $\epsilon = \|\nabla y^o\|^+ / 10$, with:

$$\|\nabla y^o\|^+ = \min\{\|\nabla y^o(\mathbf{x}, t)\|, \text{such that } \|\nabla y^o(\mathbf{x}, t)\| > 0\}. \quad (9)$$

$$\mathbf{x} \in \Omega$$

$$t \in [t_0, t_f].$$

As a consequence, this leads to the following inequality when ∇I is not null:

$$\frac{0.995}{\|\nabla y^o\|} \leq \frac{1}{\|\nabla y^o\|_\epsilon} \leq \frac{1}{\|\nabla y^o\|}$$

Therefore, $\frac{\nabla y^o}{\|\nabla y^o\|_\epsilon}$ is either null or almost equal to the unitary vector $\frac{\nabla y^o}{\|\nabla y^o\|}$. Following the notations of Section 2, the observation transformation is now defined as $\mathbf{A}_{\text{Ang}}(u) = \frac{\nabla u}{\|\nabla u\|_\epsilon}$. The corresponding functional reads:

$$J_{\text{Ang}}^o(X_0) = \sum_{t_i=t_0}^{t_f} \left\| \frac{\nabla y_{t_i}^o}{\|\nabla y_{t_i}^o\|_\epsilon} - \frac{\nabla \mathcal{H}(X_{t_i})}{\|\nabla \mathcal{H}(X_{t_i})\|_\epsilon} \right\|_{\mathbf{R}_{\text{Ang}}}^2, \quad (10)$$

where \mathbf{R}_{Ang} is the covariance matrix relative to the observations errors. As the observation transformation \mathbf{A}_{Ang} is non-linear, we need to compute the tangent linear transformation $\partial_{\mathcal{H}(X)} \mathbf{A}_{\text{Ang}}$. Denoting $q = \mathcal{H}(X)$, it can be obtained for a small perturbation γ around $\mathcal{H}(X)$ as:

$$\partial_q \mathbf{A}_{\text{Ang}} \gamma = \frac{1}{\|\nabla(q)\|_\epsilon^2} \underbrace{\begin{bmatrix} q_y^2 + \epsilon^2 & -q_x q_y \\ -q_x q_y & q_x^2 + \epsilon^2 \end{bmatrix}}_{\mathcal{Q}(q)} \nabla \gamma. \quad (11)$$

The adjoint observation transformation applied to a vector u reads:

$$(\partial_q \mathbf{A}_{\text{Ang}})^* u = -\nabla \cdot \left(\frac{1}{\|\nabla(q)\|_\epsilon^2} \begin{bmatrix} q_y^2 + \epsilon^2 & -q_x q_y \\ -q_x q_y & q_x^2 + \epsilon^2 \end{bmatrix} u \right). \quad (12)$$

¹<https://fadili.users.greyc.fr/demos/WaveRestore/downloads/mcalab/Home.html>

The right-hand side of the adjoint model is finally obtained by taking $u = \mathbf{R}_{\text{Ang}}^{-1} \left(\frac{\nabla y^o}{\|\nabla y^o\|_c} - \frac{\nabla \mathcal{H}(X)}{\|\nabla \mathcal{H}(X)\|_c} \right)$ in the previous expression.

Once again there exist an infinite number of different images y^o sharing the same gradient directions, such as the family $y_{\alpha,c}^o = \alpha y^o + c$, $\forall \alpha \in \mathbb{R}$ and $c \in \mathbb{R}$. As a consequence, the angular transformation is not able to correct the intensity of the image and will only focus on correcting the position of structures.

The eigenvalues of the symmetric matrix $Q(\mathcal{H}(X))$ are ϵ^2 and $\|\nabla \mathcal{H}(X)\|^2 + \epsilon^2$, respectively, associated to the eigenvectors $\nabla \mathcal{H}(X)$ and $\nabla^\perp \mathcal{H}(X)$. As a consequence, if $\|\nabla \mathcal{H}(X)\|$ is not zero, most of the information will be sent to $\nabla^\perp \mathcal{H}(X)$, the orthogonal direction of the current state gradient direction. On the other hand, this transformation will just realise an isotropic diffusion if $\nabla \mathcal{H}(X)$ is a null vector.

According to the current values of $\nabla(\mathcal{H}(X))^k$ and ∇y^o , the minimisation of (10) produces the following effect on $(\mathcal{H}(X))^{k+1}$:

- minimise the norm $\|\nabla(\mathcal{H}(X))^{k+1}\|$ in the areas where ∇y^o is null,
- diffuse the orientation of the observation orientation ∇y^o where $\|\nabla y^o\| > 0$ and $\nabla(\mathcal{H}(X))^k$ is null,
- maximise the scalar product $\frac{\nabla(\mathcal{H}(X))^{k+1}}{\|\nabla(\mathcal{H}(X))^{k+1}\|_c} \cdot \frac{\nabla y^o}{\|\nabla y^o\|_c}$ where ∇y^o and $\nabla(\mathcal{H}(X))^k$ are non-null.

These properties illustrates how the structure orientation of images is transferred to the model variable.

4. Error description

When dealing with real satellite data, noise can corrupt the signal from the earth's surface or the atmosphere. This noise is due to measurement uncertainties as well as perturbations involved by the vertical integration of the signal (as the atmosphere can contain clouds, rain, snow, etc.). The induced errors must be characterised in order to perform an optimal analysis.

The measurement norm (3) involved in the observation cost function (2) requires the inverse of the observation error covariance matrix \mathbf{R} . As these matrices can be huge (and even non-invertible) one needs to make some approximations.

In this part we focus on building approximate covariance matrices easily invertible in selected spaces. We would like to underline that in a Fourier, wavelet, curvelet, gradient and angular space a diagonal covariance matrix can hold information on error correlation in the pixel space (this is discussed in more detail in Section 4.3 for the wavelet case). This property has already been used in Fourier space to model homogeneous correlation in the background error

covariance matrix [for instance, in Courtier et al. (1998)] and in wavelet spaces to model heterogeneous correlations [for instance, in Fisher (2003); Deckmyn and Berre (2005); Pannekoucke (2009)].

4.1. Notations

In the following, the true error covariance matrices are denoted $\mathbf{C}_{\text{Space}}^{\text{noise}}$ depending on the *noise* type and the considered observation *Space*.

Let us first consider that the available images I^o correspond to the true state I^* corrupted with an additive time independent error η . The observation equation reads, in the pixel space, $I^o = I^* + \eta$. Assuming that the error covariance matrix, $\mathbf{C}_{\text{Pix}}^{\text{noise}}$, is known, the error covariance matrix associated to a linear transform \mathbf{A} of the observed image is:

$$\mathbf{C}_{\mathbf{A}}^{\text{noise}} = \mathbf{A} \mathbf{C}_{\text{pix}}^{\text{noise}} \mathbf{A}^T \quad (13)$$

where \mathbf{A} can be a multiscale transformation described in Section 3.2.2 or the gradient transformation from Section 3.2.3.

As these matrices can be huge (and even non-invertible) they must be approximated $\mathbf{R}_{\text{Space,Approx}}^{\text{noise}} \approx \mathbf{C}_{\text{Space}}^{\text{noise}}$. The approximations considered are:

- Scalar: $\mathbf{R}_{*,\text{Scalar}}^* = \sigma^2 \mathbf{Id}_n$,
- Diagonal: $\mathbf{R}_{*,\text{Diag}}^* = \text{Diag}(\mathbf{C}_*)$,
- Block diagonal: $\mathbf{R}_{*,\text{Block}}^*$, see e.g. Appendix A for more details,

where \mathbf{Id}_n is the $n \times n$ identity matrix. For instance, when considering an independent and identically distributed (iid) noise, the matrix $\mathbf{R}_{\nabla, \text{Block}}^{\text{id}}$ is a block diagonal approximation of the true error covariance matrix $\mathbf{C}_{\nabla}^{\text{id}}$ for the gradient space ∇ . Throughout this paper, we use these denominations (scalar, diagonal, block diagonal) to refer to the above approximations, in particular ‘scalar’ means ‘proportional to identity’.

4.2. Independent additive noise

Assuming that images are corrupted by a Gaussian white noise independent and identically distributed in space and time, let us describe the error covariance matrices built in this case for the various image comparisons introduced in Sections 3.2.1, 3.2.2, 3.2.3 and 3.2.4.

Pixel case:

The true error covariance matrix $\mathbf{C}_{\text{Pix}}^{\text{id}}$ is scalar. As a consequence, there is no need to approximate this matrix:

$$\mathbf{R}_{\text{Pix,Scalar}}^{\text{id}} = \mathbf{C}_{\text{Pix}}^{\text{id}} = \sigma^2 \mathbf{Id}_n, \quad (14)$$

where the variance σ^2 monitors the noise amplitude.

Fourier case:

Here again, the true covariance matrix \mathbf{C}_{Fou}^{iid} is scalar. As a consequence, there is no need to approximate this matrix:

$$\mathbf{R}_{Fou,Scalar}^{iid} = \mathbf{C}_{Fou}^{iid} = \sigma^2 \mathbf{Id}_n, \quad (15)$$

where σ^2 monitors the noise amplitude.

Wavelet case:

The error covariance matrix in a wavelet space is:

$$\mathbf{C}_W^{iid} = \sigma^2 \mathbf{W} \mathbf{W}^T$$

where \mathbf{W} is the wavelet transform defined in Section 3.2.2. As only orthonormal wavelets transforms are used in our study, we have:

$$\mathbf{W}^T = \mathbf{W}^{-1}.$$

Then, whatever the chosen wavelet basis, provided that it is orthonormal, the error covariance matrix is

$$\mathbf{R}_{W,Scalar}^{iid} = \mathbf{C}_W^{iid} = \sigma^2 \mathbf{Id}_n. \quad (16)$$

Curvelet case:

The curvelet transform by wrapping (denoted \mathbf{C}), defined in Candes et al. (2006), is a redundant (and therefore non-invertible) transform. A pseudo-inverse of the curvelet transform by wrapping is the adjoint of the forward transform (\mathbf{C}^T). It enables a perfect reconstruction of the original image given the full set of curvelet coefficients.

The error covariance matrix in a curvelet space is nevertheless not invertible as it reads $\mathbf{C} \mathbf{C}_{pix}^{iid} \mathbf{C}^T = \sigma^2 \mathbf{C} \mathbf{C}^T (\neq \sigma^2 \mathbf{Id}_n)$. As a consequence, we do not try to explicitly construct it.

However, one can see that a sufficient condition to build a cost function in the curvelet space equal to the cost function in the pixel space is to choose $\mathbf{R}_{C,Scalar}^{iid} = \sigma^2 \mathbf{Id}_k$. This comes from the fact that \mathbf{C}^T is the pseudo-inverse of \mathbf{C} meaning that $\mathbf{C}^T \mathbf{C} = \mathbf{Id}_k$:

$$\begin{aligned} J_{curv}^o &= \frac{1}{\sigma^2} (\mathbf{C}(\mathcal{H}(X)) - \mathbf{C}(y^o))^T \mathbf{Id}_k (\mathbf{C}(\mathcal{H}(X)) - \mathbf{C}(y^o)) \\ &= \frac{1}{\sigma^2} (\mathcal{H}(X) - y^o)^T (\mathcal{H}(X) - y^o) = J_{pix}^o \end{aligned} \quad (17)$$

Remark 2. Keeping all of the coefficient in a multiscale family leads in this case to minimise exactly the same cost function as in the pixel case. Indeed, from eq. (7) and the fact that the considered adjoints are equal to their pseudo-inverses, we have:

$$\begin{aligned} &\sigma^2 (\mathbf{A}_m(\mathcal{H}(X)) - \mathbf{A}_m(y^o))^T \mathbf{Id}_n (\mathbf{A}_m(\mathcal{H}(X)) - \mathbf{A}_m(y^o)) \\ &= \sigma^2 (\mathcal{H}(X) - y^o)^T \mathbf{Id}_n (\mathcal{H}(X) - y^o). \end{aligned}$$

As a consequence there will be no result referring to multiscale decompositions when the observation error is uncorrelated in space.

Gradient case:

The transformation ∇ is not invertible. Neither is the error covariance matrix $\mathbf{C}_{\nabla}^{iid} = \sigma^2 \nabla \nabla^T$. In Section 5, we use two different approximations of this matrix:

- A scalar matrix corresponding to the diagonal of the true error covariance matrix:

$$\mathbf{R}_{\nabla,Scalar}^{iid} = \text{Diag}(\mathbf{C}_{\nabla}^{iid}) = \tilde{\sigma}^2 \mathbf{Id}_{2n}. \quad (18)$$

In this study, as central differences are used to compute partial derivatives, the variance reads:

$$\tilde{\sigma}^2 = \frac{\sigma^2}{2}. \quad (19)$$

- A block diagonal matrix,

$$\mathbf{R}_{\nabla,Block}^{iid} \approx \mathbf{C}_{\nabla}^{iid} \quad (20)$$

built from $\mathbf{C}_{\nabla}^{iid}$ by making the hypothesis that there is no interaction between the error made on the x - and y -derivatives. The construction and inversion of this matrix is detailed in Appendix A.

Angular case:

As the angular distance defined in Section 3.2.4 is non-linear, it is difficult to build statistics for this case. Indeed, the formal computation of the mean and the covariances of the noise error are out of reach. Therefore, we consider that observations are only reliable when $\|\nabla y^o\| \gg \sigma$. Indeed, in this case the noise does not have much impact on the gradient direction. Therefore, we design the variance of both components of $\nabla y^o(i,j) / \|\nabla y^o(i,j)\|_{\epsilon}$ to be almost equal to σ^2 for $\|\nabla y^o\| \gg \sigma$, and very large for $\|\nabla y^o\| \ll \sigma$ (so that the observation is almost discarded). This writes as follows:

$$\tilde{\sigma}_{i,j}^2 = \frac{\sigma^2}{1 - \exp\left(\frac{-\|\nabla y^o\|^2}{2\sigma^2}\right)}. \quad (21)$$

The covariance matrix used in twin experiments of Section 5 is thus:

$$\mathbf{R}_{Ang,Diag}^{iid} = \begin{pmatrix} \mathbf{R}_x & 0 \\ 0 & \mathbf{R}_y \end{pmatrix} \quad (22)$$

where we consider isotropic correlations through

$$\mathbf{R}_x = \mathbf{R}_y = \begin{pmatrix} \tilde{\sigma}_{1,1}^2 & 0 & \dots & \dots & 0 \\ 0 & \ddots & 0 & \dots & \vdots \\ \vdots & \ddots & \tilde{\sigma}_{i,j}^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & \tilde{\sigma}_{n,n}^2 \end{pmatrix}. \quad (23)$$

4.3. Correlated Gaussian white noise

Let now assume that η is an additive Gaussian noise spatially correlated. Ideally, the observation error matrices should contain off-diagonal elements to represent these correlations. However, as mentioned before, the size of such matrices for realistic application is, most of the time, far too large to be handled properly. The aim of this section is to propose computationally efficient approximations of the observation error covariances. We do so by combining the transformations \mathbf{A} with suitably chosen diagonal matrices \mathbf{R} .

Pixel case:

In these experiments, the noise distribution is chosen such that each pixel variance is the same. Therefore, in the pixel space the approximation is scalar and reads:

$$\mathbf{R}_{Pix,Scalar}^{cor} = \text{Diag}(\mathbf{C}_{Pix}^{cor}) = \sigma^2 \mathbf{Id}_n. \quad (24)$$

Wavelet case:

- A naive approach would be to build the error covariance matrix in wavelet space from the diagonal assumption in the pixel space given by eq. (24). Doing this leads to define the approximation of the covariance matrix in any orthonormal wavelet subspace as:

$$\begin{aligned} \mathbf{R}_{\mathbf{W},Scalar}^{cor} &= \mathbf{W} \text{Diag}(\mathbf{C}_{pix}^{cor}) \mathbf{W}^T = \sigma^2 \mathbf{Id}_n \\ &\neq \text{Diag}(\mathbf{C}_{\mathbf{W}}^{cor}). \end{aligned} \quad (25)$$

As presented for the uncorrelated case in remark 2, the cost function in the pixel space involving $\mathbf{R}_{Pix,Scalar}^{cor}$ matrix is equal to the one in the wavelet space involving $\mathbf{R}_{\mathbf{W},Scalar}^{cor}$. Therefore, in this work, no results will be presented with this approximation.

- One can also build an approximation covariance matrix from the true covariance matrix in a wavelet space $\mathbf{C}_{\mathbf{W}}^{cor}$. The latter can be constructed from \mathbf{C}_{pix}^{cor} using eq. (13). In Vannucci and Corradi (1999), the authors provide an algorithm to do so for a 1-D wavelet basis. We developed a 2-D version of this algorithm to perform the experiments presented in Section 5.4.

In this study, we approximate the true error covariance matrix $\mathbf{C}_{\mathbf{W}}^{cor}$ by its diagonal:

$$\mathbf{R}_{\mathbf{W},Diag}^{cor} = \text{Diag}(\mathbf{C}_{\mathbf{W}}^{cor}) = \text{Diag}(\mathbf{W} \mathbf{C}_{pix}^{cor} \mathbf{W}^T). \quad (26)$$

Such a process is applied with the Haar basis and the Daubechies basis with eight vanishing moments. In the following, their covariances matrix are denoted as:

$$\mathbf{R}_{D8,Diag}^{cor} \text{ for the Daubechies basis} \quad (27)$$

$$\mathbf{R}_{Haar,Diag}^{cor} \text{ for the Haar basis} \quad (28)$$

Remark 3. As the transformation $\mathbf{A}_m = \mathbf{W}$ is bijective, the information used by the assimilation process is the same as in the pixel comparison. Therefore, the only difference between (6) and (7) lies in the error statistics description through $\mathbf{R}_{D8,Diag}^{cor}$ or $\mathbf{R}_{Haar,Diag}^{cor}$, coupled with the transformations $\mathbf{W} = D8$ or $\mathbf{W} = Haar$.

Curvelet and Fourier case:

In these cases, we do not build the true covariance matrix \mathbf{C}_{Curv}^{cor} and \mathbf{C}_{Fou}^{cor} . We rather estimate the diagonal of each matrix (variances) from $r = 1000$ independent noise realisations.

The variance of a given (Fourier or curvelet) coefficient y is estimated by:

$$\text{var}(y) = \frac{1}{r-1} \sum_{i=1}^r (y_i - \hat{y})^2. \quad (29)$$

Gradient case:

When the noise is identically distributed and the correlations are isotropic, the variance of each gradient component is the same. Therefore, taking into account information exclusively about variances leads to work with the scalar matrix:

$$\mathbf{R}_{\nabla,Scalar}^{cor} = \text{Diag}(\mathbf{C}_{\nabla}^{cor}) = \tilde{\sigma}^2 \mathbf{Id}_{2n}. \quad (30)$$

Note that, as central differences are used to compute partial derivatives, the variance reads:

$$\tilde{\sigma}^2 = \sigma^2 \frac{1-p}{2} \quad (31)$$

where

$$\begin{aligned} p &= \text{cov}(y^o(i+1, j), y^o(i-1, j)) \\ &= \text{cov}(y^o(i, j+1), y^o(i, j-1)). \end{aligned} \quad (32)$$

Note that when p is close to 1, $\tilde{\sigma}^2$ is far smaller than σ^2 , so the noise in gradient space is smaller than that of pixel space.

Angular case:

As it is very complex to specify the effect of a correlated noise for this transformation, we simply use the covariance matrix built for uncorrelated noise:

$$\mathbf{R}_{Ang,Diag}^{cor} = \mathbf{R}_{Ang,Diag}^{iid}$$

4.3.1. Illustration. Let us describe what kind of correlations in the pixel space are taken into account in the cost function with the proposed diagonal matrices in wavelet, curvelet and Fourier spaces. To do so, we build a cost function in the pixel space (6) equivalent to the one in a multiscale space (7) by creating a new covariance matrix named $\mathbf{R}_{Pix,Eq}^{cor}$. Indeed, eq. (7) reads:

$$(\mathcal{H}(X) - y^o)^T \mathbf{A}_m^T (\mathbf{R}_{\mathbf{A}_m,Diag}^{cor})^{-1} \mathbf{A}_m (\mathcal{H}(X) - y^o) \quad (33)$$

This equation is equal to:

$$(\mathcal{H}(X) - y^o)^T (\mathbf{R}_{Pix,Eq}^{cor})^{-1} (\mathcal{H}(X) - y^o). \quad (34)$$

where

$$(\mathbf{R}_{Pix,Eq}^{cor})^{-1} = \mathbf{A}_m^T (\mathbf{R}_{A_m,Diag}^{cor})^{-1} \mathbf{A}_m. \quad (35)$$

Therefore, the covariance matrix $\mathbf{R}_{Pix,Eq}^{cor}$ used in the pixel space enables to build a cost function equal to the one involving the transform \mathbf{A}_m in association with the diagonal matrix $\mathbf{R}_{A_m,Diag}^{cor}$. Thanks to this equivalent matrix in the pixel space $\mathbf{R}_{Pix,Eq}^{cor}$ it is easy to visualise correlation: Fig. 2 exhibits error correlation between a pixel and its neighbourhood at nine different locations.

The top left panel presents the true error correlations extracted from \mathbf{C}_{pix}^{cor} . The correlations are isotropic and the same for each pixel *far* from the boundary. The other panels present the equivalent correlations in the pixel space induced by a diagonal matrix in the studied spaces, using $\mathbf{R}_{Pix,Eq}^{cor}$. The correlations modelled using the Fourier, curvelet or Daubechies transforms look similar to the true ones, despite being built using diagonal matrices.

However, one can notice that:

- the isotropy and homogeneity of the noise is not represented in Daubechies Wavelet basis,
- some spurious correlations do appear in the Curvelet and Daubechies spaces,
- unlike the true correlation map, the correlation map induced by periodic transforms (Fourier, Daubechies Wavelet, Curvelet) is periodic.

The impact of accounting for those correlations along the assimilation process is studied in Section 5.4.

Remark 4. The same comparison is difficult to set up with gradient and angular transform because they induce a loss of information. Therefore, no trivial comparison between the correlations taken into account for *pixel gradients/angular* can be made.

5. Twin experiments and robustness to noise

The aim of this section is to investigate the performance of the proposed sparse representation of observation errors statistics (combined with the proposed transformations) depending on noise characteristics. First the experimental framework is presented, and the transformations are validated through a noise-free experiment. Note that from now on, we will use the word ‘transformation’ also for the pixel comparison, for which the transformation is simply $\mathbf{A} = \mathbf{Id}_n$. Then three kinds of observation noise are studied: an uncorrelated and two different spatially correlated Gaussian noises. The relative merits of the studied observation transformations \mathbf{A} combined with their associated \mathbf{R}_A matrix approximations are discussed.

5.1. Experimental framework and dynamical model

The experimental framework mimics the drift of a vortex on a turntable. The evolution of a vortex in the atmosphere is simulated at the CORIOLIS experimental

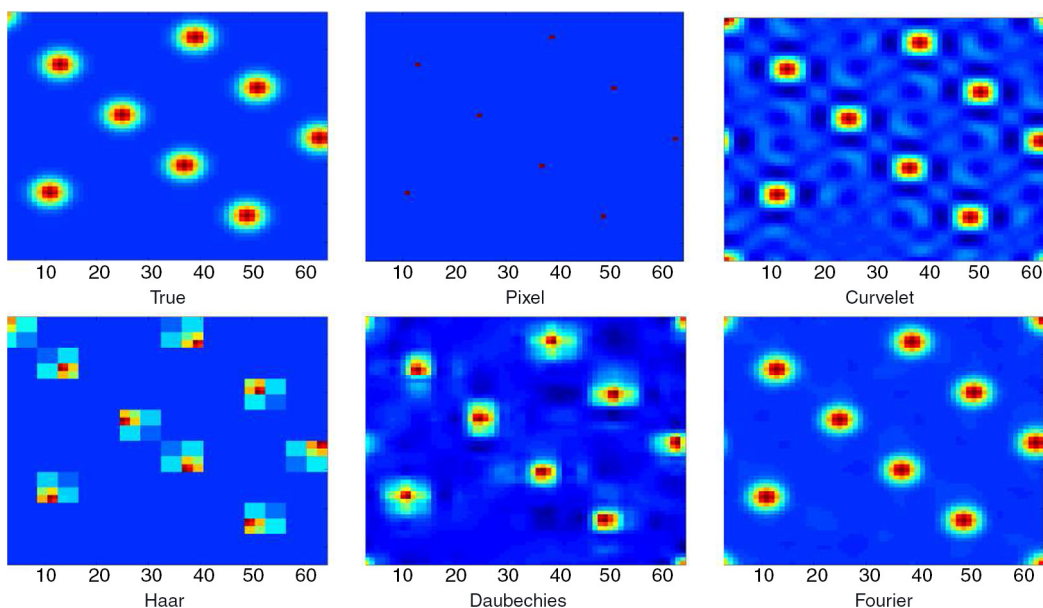


Fig. 2. Spatial correlations around nine selected points (top-left) and their representation with a diagonal approximation of the \mathbf{R} matrix combined with various image transformations.

turntable² (Grenoble, France) which re-creates the effect of the Coriolis force on a thin layer of water. A complete rotation of the tank takes 60 seconds which corresponds to one Earth rotation. The vortex is created by stirring the water and made visible thanks to the addition of a passive tracer (fluorescein). The camera is placed above the turntable, and photographs of the vortex constitute the observed image sequence. For more details about these experiments, see Flór and Eames (2002).

5.1.1. Numerical configuration. In this configuration, the evolution of the fluid can be represented by the shallow-water equations involving the horizontal velocity $\mathbf{w}(\mathbf{x}, t) = (u(\mathbf{x}, t), v(\mathbf{x}, t))$, where u and v are the zonal and meridional components of the velocity, and the water elevation $h(\mathbf{x}, t)$. These unknown variables are defined on the spatial domain $\Omega \ni \mathbf{x}$ and the time interval $[t_0, t_f] \ni t$. Such a model reads:

$$\begin{cases} \partial_t u - (f + \zeta)v + \partial_x B &= -ru + \kappa \Delta u \\ \partial_t v + (f + \zeta)u + \partial_y B &= -rv + \kappa \Delta v \\ \partial_t h + \partial_x(hu) + \partial_y(hv) &= 0. \end{cases} \quad (36)$$

The relative vorticity is denoted by $\zeta = \partial_x v - \partial_y u$ and the Bernoulli potential by $B = g^* h + \frac{u^2 + v^2}{2}$, where g^* is the reduced gravity. The Coriolis parameter on the β -plane is given by $f = f_0 + \beta y$, κ is the diffusion coefficient and r the bottom friction coefficient. The following numerical values were used for the experiments: $r = 9.10^{-7}$, $\kappa = 0$, $f_0 = 0.25$, $g = 9.81$ and $\beta = 0.0406$.

The simulation is performed on a rectangular domain $[0, L] \times [0, H]$ representing a sub-domain of the turntable with $L = H = 2.525$ m. The domain is discretised on a 128×128 uniform Arakawa C-type square grid. A finite differences scheme is used for space discretisation. Time integration is performed using a fourth order Runge-Kutta scheme. The time step of the turntable experiment is set to 0.01s which would correspond to 14.4s in the atmosphere.

5.1.2. Observation operator. The vortex temporal evolution is shown through the fluorescein concentration evolution. This evolution is observed by an image sequence whose grey levels correspond to the concentration of a passive tracer q transported by the velocity field.

Denoting \mathbf{w} the velocity (computed by the model \mathcal{M}) transporting the passive tracer and ν_T the diffusion coefficient, we have

$$\begin{cases} \partial_t q + \nabla q \cdot \mathbf{w} - \nu_T \Delta q = 0 \\ q(t_0) = q_0. \end{cases} \quad (37)$$

Assuming that the initial concentration of q is known at time t_0 , the dynamic of q on the time interval $[t_0, t_f]$ is defined by the model (37), where the diffusion coefficient is $\nu_T = 10^{-5}$.

In the following experiments the considered image observation sequence y^o represents observations of q . As a consequence, the observation operator reads:

$$\mathcal{H}(X_{t_i}) = q(t_i). \quad (38)$$

where $q(t_i)$ comes from (37).

Remark 5. In those experiments, we assume that the initial concentration of the passive tracer is known. Therefore, we do not control q_0 .

5.1.3. Twin experiments configuration. In order to focus on the methodological aspects we will use a so-called twin experiment framework. In this classical approach, synthetic observations are created thanks to a model simulation from a known ‘true state’; then an assimilation experiment is performed starting from another ‘background’ state using the synthetic observations. The result of this analysis can be compared with the synthetic truth.

In our case, considering initial conditions at time t_0 of the state variables (u_0^* , v_0^* and h_0^*) and fluorescein concentration q_0^* , a simulation of the dynamical model is made from $t_0 = 0s$ to $t_f = 6s$. This defines a scenario consistent with realistic experiments, as illustrated in Fig. 3.

The simulated values of the concentration $q^*(t_i)$, at various observation times $t_i \in]t_0, t_f[$, are then used as observations. One such observation is taken every 0.25s, so that we get 25 observed images per assimilation window.

In the experiments, a variational data assimilation algorithm, allowing the estimation of the state variable (u, v, h) , is applied. The background variables $u^b(t_0, \mathbf{x})$, $v^b(t_0, \mathbf{x})$ are initialized at rest while $h^b(t_0, \mathbf{x}) = h_{mean} \forall \mathbf{x}$ is initialized as a constant. The minimisation of the cost function is performed thanks to N1QN3 solver (see Gilbert and Lemaréchal, 1989).

Notations:

The true state variables and observations are denoted with the superscript $*$, the background state (without assimilation) with b , and the analysis (the estimated state after assimilation) with a .

5.1.4. Metrics for qualitative analysis. Here we introduce some tools in order to measure different criteria on the estimated velocity fields. We mainly focus on the estimation errors for the velocities u and v . The vorticity error and the angular error are also looked at, as it gives an excellent criteria to evaluate the quality of the estimated velocities directions.

²<http://coriolis.legi.grenoble-inp.fr>

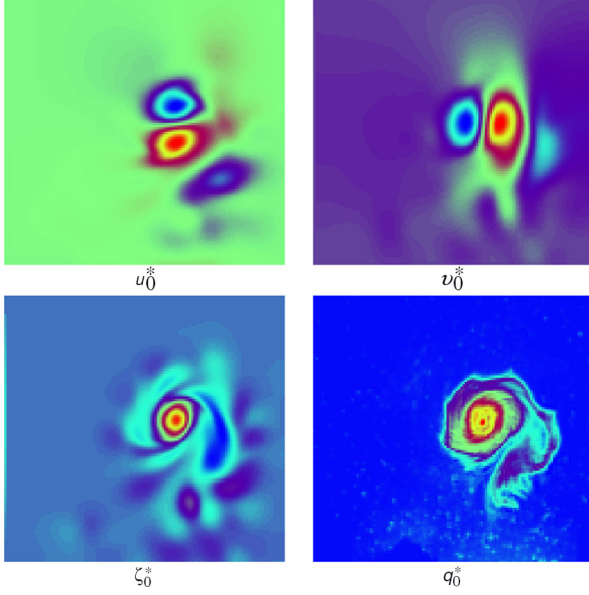


Fig. 3. Initial values of the scenario chosen for the twin experiments. The zonal and meridional velocities u_0^* and v_0^* are characterised by a strong vortex, illustrated by the initial vorticity ζ_0 . The height h_0^* is assumed almost flat. The velocities take their values from -0.03 ms^{-1} (blue) to 0.04 ms^{-1} (red) and the 2-D vorticity from -0.2 s^{-1} (blue) to 0.64 s^{-1} (red). This synthetic initialisation has been created in order to match correctly the initialisation of the passive tracer, q_0^* .

In the following, we refer to the Root Mean Square Error (RMSE) between the variables u (either u^b for the background without assimilation or u^a for the analysis with assimilation) and the true synthetic ones u^* at time t_0 :

$$RMSE(u) = \sqrt{\frac{1}{|\Omega|} \sum_{\mathbf{x} \in \Omega} (u(t_0, \mathbf{x}) - u^*(\mathbf{x}))^2}.$$

Similarly, we define the RMSE for v and the vorticity $\zeta = v_x - u_y$, as well as the angle α defined as:

$$\alpha = \arccos\left(\frac{\langle U_\epsilon, U_\epsilon^* \rangle}{\|U_\epsilon\| \|U_\epsilon^*\|}\right), \quad (39)$$

where $U_\epsilon(\mathbf{x}) = [u(t_0, \mathbf{x}), v(t_0, \mathbf{x}), \epsilon]$, $U_\epsilon^*(\mathbf{x}) = [u^*(t_0, \mathbf{x}), v^*(t_0, \mathbf{x}), \epsilon]$ and $\|U_\epsilon\|^2 = \langle U_\epsilon, U_\epsilon \rangle$. This corresponds to a modification of the angular error of Barron et al. (1994) as proposed in Souopgui (2010).

In the following, we mostly consider the ratio $RMSE(u^a)/RMSE(u^b)$. This ratio represents the residual error of the analysed zonal velocity u^a with respect to the error of the background zonal velocity u^b . A ratio smaller than 1 means that the assimilation has improved the corresponding variable with respect to the background.

Remark 6. In our experiments, as the background velocities are initialised with null values, this ratio reads:

$$\frac{\sqrt{\sum_{\mathbf{x} \in \Omega} (u^a(t_0, \mathbf{x}) - u^*(t_0, \mathbf{x}))^2}}{\sqrt{\sum_{\mathbf{x} \in \Omega} (u^*(t_0, \mathbf{x}))^2}} \quad (40)$$

Therefore, the square of this number represents the percentage of noise (in terms of energy) present in the analysed field.

5.2. Validation with perfect data

We want to compare the behaviour of *gradients* (see Section 3.2.3) and *angular* (see Section 3.2.4) to *pixel*, *curvelet* and *wavelet* observation transformations (see Section 3.2) in the case of perfect data (i.e. without any noise).

As we started with a static background, we have no confidence in it. Therefore, we introduce a weight w_b in the cost function:

$$J = J^o + w_b J^b$$

Choosing w_b that is small ensures that the J^b term acts only as a regularisation term, and not as a strong feedback to the background.

Figure 4 shows the evolution of the RMSE ratio of the zonal velocity u along the iterations for *pixel*, *gradients* and *angular*. For each distance notion, this ratio decreases with the iterations. The minimisation leads to a coherent analysed field in each case. The best zonal velocities field is reconstructed with the *angular* observation transformation.

RMSE ratios obtained at the end of the minimisation processes are presented in Table 1.

As illustrated by Fig. 4, most of the initial velocity error is corrected throughout data assimilation. Indeed, for the velocities components (u and v), the noise energy of the analysed field represent less than 2.5% of the true field energy.

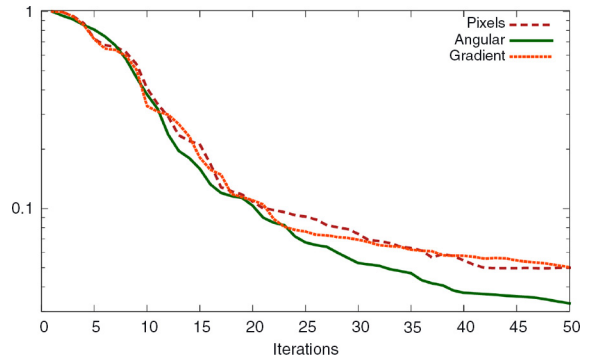


Fig. 4. Evolution of the ratio of RMS errors of the velocity u with respect to 4-D-Var iterations. Perfect data are observed and the three transformations are compared.

Table 1. Ratio between analysed state RMS and background state RMS for perfect data

| | $\frac{\text{RMSE}(y^a)}{\text{RMSE}(y^p)}$ | $\frac{\text{RMSE}(y^r)}{\text{RMSE}(y^p)}$ | $\frac{\text{RMSE}(z^a)}{\text{RMSE}(z^p)}$ | $\frac{\text{RMSE}(z^r)}{\text{RMSE}(z^p)}$ |
|----------|---|---|---|---|
| Pixels | 4.7% | 3.5% | 11.4% | 24.5% |
| Angular | 3.6% | 2.5% | 9.1% | 21.5% |
| Gradient | 4.8% | 3.5% | 11.0% | 23.3% |

This ratio is computed after assimilation starting from a background at rest.

One can notice that the *angular* observation transformation leads to slightly better results than other observation transformations for all criteria. Therefore, the loss of information induced by the *angular* transformation is not a problem in a perfect data context.

In this ‘perfect observation’ case, each observation transformation gives satisfying results. In particular, it validates the proposed gradient and angular transformations.

In realistic applications, there is no such thing as perfect observations. The next two sections study the impact of data noise on the assimilation process. In Section 5.3, we perform experiments with a Gaussian space- and time- uncorrelated noise, and then in Section 5.4 the impact of spatial correlation is discussed. For each kind of noise, different noise levels are tested. For each level, 10 noise scenarios are generated in order to perform 10 independent experiments.

In order to quantify the reliability of the information available in our image sequences, we refer to the Signal to Noise Ratio (SNR) which is computed as:

$$\text{SNR} = 10 \log_{10} \left(\frac{\sum_{\mathbf{x} \in \Omega} (y^*(\mathbf{x}))^2}{\sum_{\mathbf{x} \in \Omega} (y^*(\mathbf{x}) - y^o(\mathbf{x}))^2} \right) \quad (41)$$

which represents the logarithm of the ratio between Image Energy and Noise Energy, where y^* represents the true synthetic image and y^o the noisy image. A large SNR ($\approx 30 - 40\text{dB}$) means that the image sequence has a good quality whereas a smaller (or negative) SNR means that the image sequence is strongly corrupted by noise. Figure 5 presents examples of noisy images used in our experiments.

5.3. Uncorrelated additive Gaussian white noise

We first consider images corrupted by an additive Gaussian white noise uncorrelated in space and time. The noise is identically distributed with mean 0 and variance σ^2 . We test the robustness of our transformations to different noise scenarios, considering $\sigma \in [0; 0.1]$ (whereas $y^* \in [0.13; 0.86]$).

Table 2 presents the ratio between analysed state RMSE and background state RMSE for several noise perturbations. One can see that the *angular* transformation is not well suited to deal with independent Gaussian white noise.

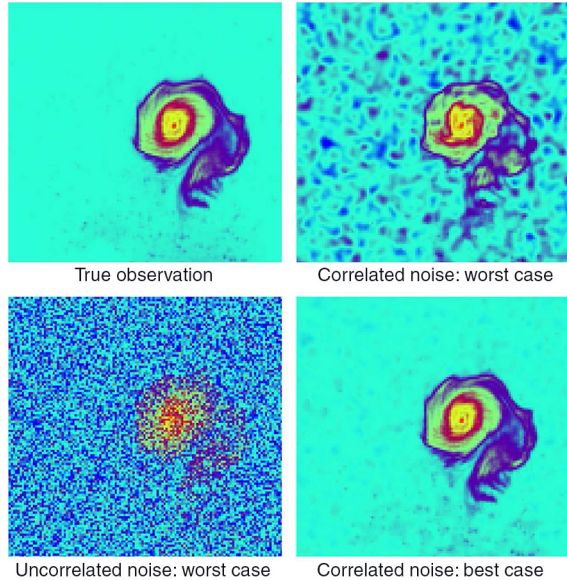


Fig. 5. First image of the sequence used for the assimilation, for various noise levels. The tracer concentrations vary from 0 (blue) to 1 (red). Top left: Perfect data scenario. Top right: Worst scenario tested for additive correlated Gaussian white noise (SNR = 14.8 dB). Bottom left: Worst scenario for Gaussian white noise (SNR = 6.7 dB). Bottom right: Best scenario for an additive correlated white noise (SNR = 26.8 dB).

Indeed, even with images sequences of good quality, the error on the analysed field remains important. This can be explained by the fact that the normalised gradient field holds less information than other distances. This is particularly true in flat areas where the information is not reliable, as taken into account by the covariance matrix modelling (21–23). On top of that, the angular covariance matrix modelling suffers from the fact that no correlations are considered. All those drawbacks could explain why the result obtained with this distance are worse than the results obtained with the other distances (whereas they were the best with perfect data).

Results obtained with the block covariance matrix, $\mathbf{R}_{\nabla, \text{Block}}^{\text{id}}$ of eq. (20), or the scalar matrix, $\mathbf{R}_{\nabla, \text{Scalar}}^{\text{id}}$ of eq. (18) are very different. In this context, introducing some information on the correlation through the covariance matrix leads to a better analysed field for all high noise levels.

To summarise, in case of additive uncorrelated Gaussian white noise, all the observation transformations are sensitive to the noise amplitude. In this particular case, we cannot conclude that working in a space is better than working in another one. Indeed, results obtained for the *pixel* or even *gradients* transformations (coupled with adequate covariance matrices) are very close to each other. In this case, the use of a diagonal \mathbf{R} matrix in the pixel space is not an approximation; therefore, it is difficult to improve the results by using different transformations. However, this

Table 2. Ratio between analysed state RMSE and background state RMSE, for the independent additive noise scenarios (Section 5.3)

| SNR | High noise | | | Small noise |
|-------------------|------------|---------|---------|-------------|
| | 6.8 dB | 12.3 dB | 26.2 dB | 46.3 dB |
| Pix – Scalar | 61.7% | 32.0% | 11.1% | 5.3% |
| ∇ – Block | 62.7% | 28.7% | 10.8% | 5.1% |
| ∇ – Scalar | Divergence | 48.8% | 12.3% | 4.7% |
| Ang – Diag | 80.3% | 62.2% | 31.1% | 10.9% |

The variable of interest is u . Different noise levels are considered.

does not hold when the noise is correlated in space, as studied in the next section.

5.4. Correlated Gaussian noise: homogeneous isotropic case

Let us now consider an additive Gaussian noise spatially correlated. The correlation is modelled by:

$$\eta = G \star \beta \quad (42)$$

where β is an independent and identically distributed Gaussian white noise of variance σ , \star is the convolution product symbol and G is a Gaussian filter of size $(2n+1)(2n+1)$ such that

$$G(i,j) = \frac{1}{|G|} \exp\left(-\frac{i^2+j^2}{2\sigma_L^2}\right) \text{ for } i,j = -n, \dots, n. \quad (43)$$

where $|G| = \sum_{i,j} \exp\left(-\frac{i^2+j^2}{2\sigma_L^2}\right)$.

The correlations are isotropic and homogeneous, which means that the correlation model is exactly the same for each pixel. In our study, the isotropic correlation length of eq. (43) is parameterised with $\sigma_L = 1.5$. Figure 2 shows the true correlation and the correlation modelled with a diagonal approximation of the error covariance matrix in several spaces.

We performed various experiments, combining observation transformations, covariance matrix approximations, data thinning or compressing. To clarify the results presentation, we sorted them into two classes:

- bijective transformations (Pixels, Fourier, Wavelet and Curvelet³), with diagonal covariance matrices approximations (Section 5.4.1);
- lossy transformations (thinning, Gradient, Angular, truncated multiscale analysis), with various covariance matrices approximations (Section 5.4.2).

³Strictly speaking, the curvelet transformation is not bijective, but there exist pseudo-inverse transformations that recover the full image from its curvelets decomposition, so that it is indeed a lossless transformation.

5.4.1. *Bijective transformations.* In this paragraph the whole dataset is used, possibly after a change of variable (Pixels, Fourier, Wavelet, Curvelet).

Table 3 presents the RMSE ratio of the zonal component of the velocity with respect to the SNR.

The following remarks can be stated:

- (1) We can first see that the pixel distance combined with a diagonal matrix performs poorly, even with a small noise level, and even more so with a large one. In that case, ignoring the error correlations is severely detrimental to the analysis.
- (2) On the contrary, using a diagonal error covariance matrix in combination with a change of space re-introduces information about the noise correlation, allowing for much better results and an increased robustness to the noise level, as we can see with the curvelets, Daubechies wavelets, and Fourier.
- (3) The Haar wavelets performance is disappointing. Indeed, we saw previously (see Fig. 2) that they induced poorly shaped (non-isotropic) correlations, contrary to the other changes of basis. So using wavelets is not a sufficient condition to ensure the recovery of correlations, the choice of the wavelet basis should be made according to the observation noise correlation pattern.
- (4) Fourier performs slightly better than the others; this is not surprising since Fourier representation is well suited to represent isotropic correlation.

5.4.2. *Lossy transformations.* A common way to deal with correlated observation error is to get rid of the correlated part of the signal. This is usually done by data thinning Liu and Rabier (2002). The idea is to discard observations if they are too close, in order to make sure that the remaining observations are uncorrelated, so that the diagonal approximation may be valid in the pixel space.

In this paragraph, we investigate various ways of accounting for correlated observation errors through lossy

Table 3. Ratio between analysed state RMSE and background state RMSE for data with spatially correlated additive noise of decreasing standard deviation

| SNR | High noise | | Small noise |
|-----------------------|------------|---------|-------------|
| | 14.8 dB | 20.8 dB | 26.8 dB |
| Pix – Scalar | 36.8% | 21.8% | 15.2% |
| C – Diag | 8.3% | 7.7% | 8.1% |
| Haar – Diag | 22.5% | 12.3% | 8.2% |
| D ₈ – Diag | 9.1% | 7.5% | 7.0% |
| Fourier – Diag | 7.3% | 7.3% | 7.3% |

Bijective image transformations are applied in this case (Section 5.4.1). The variable of interest is u .

transformations. In addition to data thinning, we also present the gradient and the angular representations of the signal. Similarly, these transformations removes some part of the signal, but as it has been shown previously [see eq. (31)], it also filters the noise, provided the spatial correlation is strong enough.

Finally, we can use another way to transform the information: first we use a change of variable into wavelets or curvelets, and then we discard selectively some parts of the signal. This operation, called thresholding in the image processing community, consists of keeping only the largest coefficients of the wavelets/curvelets decomposition, and corresponds to a projection on a subspace. Strictly speaking, the aim of such an operation is not to reduce the correlation in the observation, but to compress the data.

Table 4 presents the final RMSE ratio of the zonal velocity errors (as in Table 3) for those experiments.

As before, we can draw several comments from these results:

- (1) Thinning slightly improves the results when we keep half of the pixels in each direction, meaning that indeed the undersampled observations fit the uncorrelated hypothesis better. When the thinning is too strong however [e.g. for $(1/3)^2$ thinning], it deteriorates the analysis. This can be explained by the scale of our signal: the vortex width is 20 pixels, so that the image resolution is not that high in comparison. Therefore, thinning also discards valuable information, degrading the results as a consequence.
- (2) The gradient transformation provides a better way of extracting information than pixels (while still being outperformed by wavelets/curvelets/Fourier, as shown above).
- (3) The angular transformation is disappointing in this setting. It is likely that too much information is

Table 4. Ratio between analysed state RMSE and background state RMSE for data with spatially correlated additive noise of decreasing standard deviation

| SNR | High noise | | Small noise |
|-------------------|------------|---------|-------------|
| | 14.8 dB | 20.8 dB | 26.8 dB |
| Pix – Scalar | 36.8% | 21.8% | 15.2% |
| Pix – thinning(2) | 28.8% | 19.1% | 13.2% |
| Pix – thinning(3) | 46.5% | 22.9% | 15.4% |
| ∇ – Scalar | 20.2% | 13.5% | 11.1% |
| Ang – Diag | 59.8% | 42.6% | 31.5% |
| $T_1(C)$ – Diag | 14.4% | 8.4% | 7.6% |
| $T_1(D_8)$ – Diag | 26.0% | 10.2% | 8.2% |

The variable of interest is u . Lossy image transformations are applied in this case (Section 5.4.2). *Thinning(2)* means that, in each direction, every other pixel is used. *Thinning(3)* means that, in each direction, every third pixel is used.

discarded by this transformation. Moreover, the large flat area featuring in this image sequence does not provide exploitable angular information.

- (4) Multiscale analysis with thresholding leads to improved results, compared with the other lossy transformations. However, when we compare these to Table 3 we can see that thresholding has a small negative impact on the analysis, particularly for the Daubechies wavelets. This approach may still be useful in case the dataset is too large to be handled efficiently by the data assimilation system.

5.4.3. *Rate of convergence.* Figure 6 shows the behaviour of the cost function (left) during the minimisation and the associated analysis RMS error (right) for three selected experiments (*Pix – Scalar*, *D₈ – Diag* and *C – Diag*).

The convergence rate is higher in the pixel space than in a wavelet or curvelet space (left panel). However, the minimum reached at convergence is better in the wavelet or curvelet space (right panel).

For those experiments, the RMSE decrease rate is the same in the pixel space and wavelet space. For *C – Diag*, incorporating information on error correlation leads to a degradation of the RMSE decrease rate. This may be due to the difficulty to properly estimate the diagonal elements of \mathbf{R} .

For real applications this can be a problem as it means that if the assimilation process is stopped too soon, using curvelet signal representation can be worse than using the pixel representation.

5.4.4 *Error analysis.* Figure 7 presents the analysis error on v_0 for different observation transformations. The results are presented for the scenario involving the worst noise level studied in a correlated case (see Fig. 5 and Sections 5.4.1 and 5.4.2). This confirms the good performance of Fourier, Curvelet, Daubechies. One can add the following remarks:

- Fourier, Curvelet, Daubechies analysis errors are mostly located near the vortex area. Indeed, the analysed velocity field is too smooth to represent correctly the singularities around the vortex.
- Pixel, gradient and angular analysis errors are not limited to the vortex area. This suggests that it is hard to distinguish the error from the signal when poor information about the error structure is provided.
- The analysis error when using Haar basis is important on the whole domain. This is another illustration of the (relative) shortcoming of the diagonal hypothesis in this basis for this kind of noise.

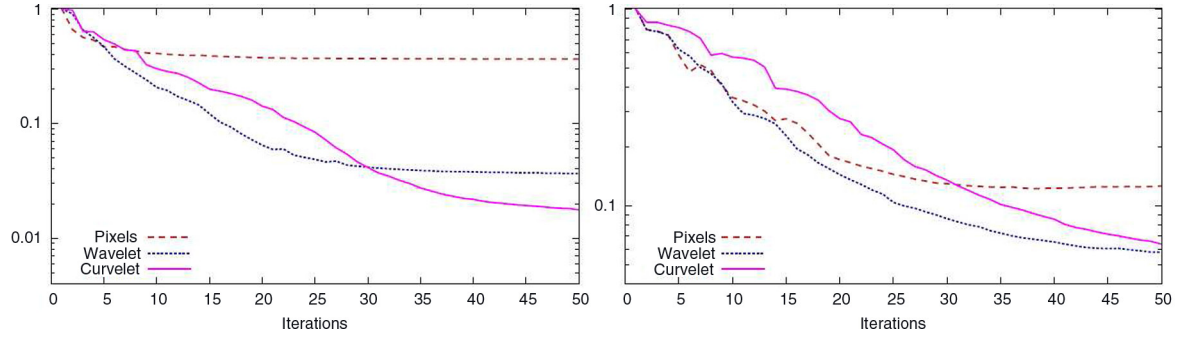


Fig. 6. Mean (over 10 experiments) cost function evolution with respect to 4-D-Var iterations (left). Mean (over 10 experiments) evolution of the RMSE ratio of v component of the velocity with respect to 4-D-Var iterations (right). Convergence rates are plotted for an isotropic noise. The noise magnitude is the smallest considered in the various experiments (26.8 dB).

5.4.5. *Summary.* In this section we showed that it is possible to account for homogeneous and isotropic observation noise using various data transformations coupled with diagonal approximations of the error correlation matrices.

Indeed the diagonal matrix associated to the Fourier (resp. Curvelet or Daubechies Wavelet) implicitly takes into account the noise spatial correlations, without requiring computationally expensive matrix inversions.

Simpler solutions have also been presented. Among those solutions, considering the signal in a gradient space seems to be an easier yet valid approach.

5.5. Correlated Gaussian white noise: an anisotropic and inhomogeneous case

Multiscale transformations coupled with non-constant diagonal error correlation matrix approximations gave

good results for isotropic correlated noise. The Curvelet transform did not bring improvement over Fourier and Wavelet, despite being more complex and expensive. This is not surprising since a homogeneous and isotropic noise can be represented by a diagonal matrix in Fourier space. Therefore, on the previous case, nothing can be gained using other transform since Fourier representation (associate to a diagonal matrix) is optimal. However, curvelets provide a flexible way of dealing with more generic noise structures. In this section, we consider anisotropic and inhomogeneous Gaussian noise.

Figure 8 presents an example of noise realisation, as well as its impact on a tracer observation. The main direction as well as the length scale of the noise correlation function differ from one pixel to another.

Table 5 presents the RMSE ratios at the end of the analysis process. As for the isotropic case, it is better to

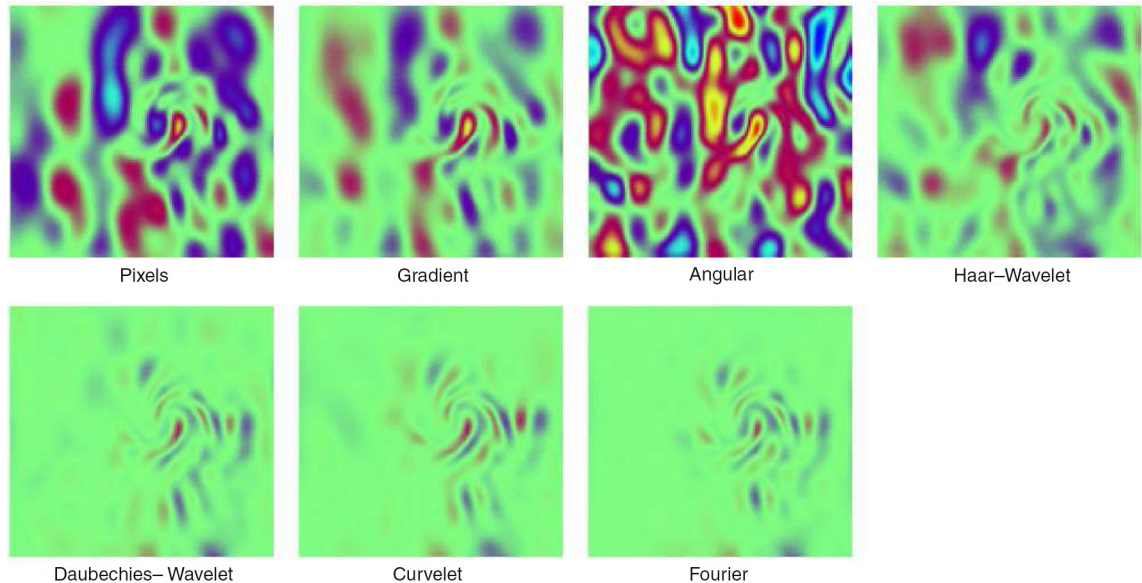


Fig. 7. Error analysis on the v -component of the velocity using the worst correlated image sequence studied. The velocity errors range from -0.0075 ms^{-1} (blue) to 0.0075 ms^{-1} . The true velocity field (see Fig. 3) ranges from -0.025 ms^{-1} to 0.0405 ms^{-1} .

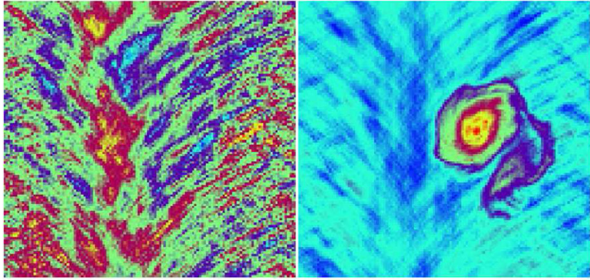


Fig. 8. Left: An example of anisotropic inhomogeneous Gaussian white noise. Right: A corrupted observation.

take into account an approximation of the error correlation and whatever the noise level is. Indeed, when the signal is considered in the Daubechies Wavelet basis, in the Fourier space or in the Curvelet space the error at the end of the assimilation is considerably reduced (compared to pixel case).

However, unlike in the isotropic case, Curvelet representation outperforms Wavelet and Fourier. Indeed, each atom of the curvelet family has three characteristics: localisation, scale, orientation. The latter allows for better anisotropic representation.

6. Conclusion

This paper first presents two original observation transformations dedicated to image-type observation assimilation. Both are based on the gradient of the image, in norm and orientation, respectively. They showed competitive performances on an academic test case compared to previously introduced metrics and therefore are good candidates for more realistic applications. In the future, it could be interesting to study combinations of such metrics in order to benefit from their respective merits.

The second and main part of this paper highlights the importance of a good specification of the observation error covariance matrix in data assimilation. Indeed, this is a crucial point for image-type data, but is likely to be of

Table 5. Ratio between analysed state RMSE and background state RMSE for data with spatially correlated additive noise of decreasing standard deviation

| SNR | High noise | | Small noise |
|-----------------------|------------|---------|-------------|
| | 12.5 dB | 18.5 dB | 26.5 dB |
| Fourier – Diag | 23.5% | 13.8% | 8.2% |
| C – Diag | 14.0% | 9.7% | 8.1% |
| D ₈ – Diag | 24.3% | 14.6% | 8.3% |
| Pix – Scalar | 47.2% | 35.7% | 16.1% |

In this case an anisotropic and inhomogeneous noise (Section 5.5, see Fig. 8) is applied to the observations. The variable of interest is u .

significant importance for other kinds of data. This was illustrated thanks to the study of two kinds of observation noise. First an uncorrelated additive Gaussian noise, for which all operators (classical and new) show similar robustness (except for the *angular* operator which suffers from the variance specification) to the noise level, and none of them seem any better or worse. This is a favourable case that is seldom encountered in practice. More realistically, a second set of experiments is performed applying a correlated additive Gaussian noise to the data. In that case, it appears clearly that it is important to improve the observation error covariance matrix specification and avoid the use of scalar ones (i.e. $\mathbf{R} = \sigma^2 \mathbf{Id}$), as it is commonly done in practice. Interestingly, such a process can be done at a cheap cost by combining a variable transformation with a non-constant diagonal approximation for the covariance matrix. Encouraging results are obtained with Fourier and wavelet transforms for the isotropic noise and with curvelet transform in the anisotropic case. It is also pointed out that the specification of the covariance matrices should obviously be related to the actual observation noise and that it is not just a matter of using a non-diagonal \mathbf{R} .

This last remark means that a careful study of the observation error has to be done prior to assimilation. In this paper, we assumed here that \mathbf{R} is known in the pixels space, but in practice, how could we adapt this work to cases where \mathbf{R} should be modelled from scratch? On that point, a wide literature exists in different fields such as images and statistics which aims to estimate and/or model the variance terms in a different space, we can for example mention the interesting work of Nguyen van Yen et al. (2012) which iteratively estimate the wavelet variance scale by scale. In variational data assimilation, the balance between the background and observation error matrices \mathbf{B} and \mathbf{R} is a subtle alchemy, improving the \mathbf{R} description should allow for a more consistent \mathbf{B} specification.

The solutions proposed in this paper are dedicated to image observation or more generally to dense observations that can be considered to be on a given grid. It is natural to extend the proposed solution for errors correlated in time (image sequences, for instance), using 3-D transformations (2-D+time). It can also be extended to sparser and randomly spaced data but the definition of the appropriate transformation becomes difficult and will change from one observation time to the next.

Another crucial point to sort out prior to operational use, specific to image-type observation, is to address the problem of occultation (a cloud passing by over ocean surface data, for instance) and its impact on error statistics. Again, multiscale description, or combination of metrics presented in this paper, may offer a suitable framework for dealing with such problem. Some preliminary results can be found in Chabot (2014).

7. Acknowledgements

This work got support from the LEFE program of INSU (CNRS) and from the French national research agency project Geo-FLUIDS no. ANR-09-COSI-005. The authors are also grateful to the two anonymous reviewers whose comments helped to improve the manuscript.

8. Appendix A

Covariance matrix for gradient additive noise

A1 Block approximation

Recalling that the observation operators reads: $\nabla y^o = \nabla \mathcal{H}(X) + \nabla \eta$. A possible discrete implementation of the gradient $\nabla \eta = [\eta_x, \eta_y]^T$ at pixel (i, j) with $i = 1 \dots L$ and $j = 1 \dots H$, is defined as:

$$\eta_x(i, j) = \frac{\eta(i, j+1) - \eta(i, j-1)}{2} \text{ if } 1 < j < H \quad (\text{A1})$$

and

$$\eta_y(i, j) = \frac{\eta(i+1, j) - \eta(i-1, j)}{2} \text{ if } 1 < i < L \quad (\text{A2})$$

Even if the noise is a white Gaussian additive noise, there are some correlation between the gradient components. Indeed, when $\eta_x(i, j)$ is inside the domain boundary, it is correlated with $\eta_x(i+2, j)$, $\eta_x(i-2, j)$, $\eta_x(i-1, j-1)$, $\eta_x(i+1, j-1)$, $\eta_x(i-1, j+1)$, $\eta_x(i, j-1)$.

As illustrated in Fig. 9 the covariance coefficients are:

$$\begin{aligned} \text{cov}(\eta_x(i, j), \eta_y(i-1, j+1)) &= \text{cov}(\eta_x(i, j), \eta_y(i+1, j-1)) \\ &= \frac{\sigma^2}{4} \end{aligned}$$

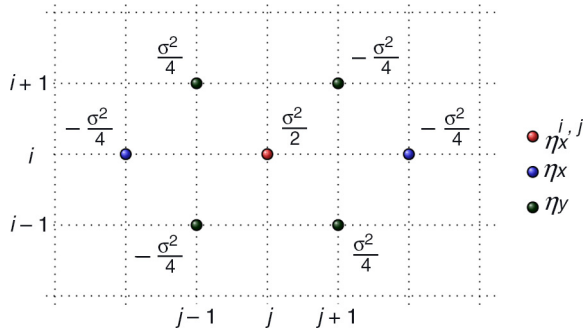


Fig. 9. Covariance between $\eta_x(i, j)$ (in red) and other elements (in green and blue) in a gradient space for an independent and identically distributed Gaussian white noise in the pixel space. The blue (resp. the green) points represents x - (resp. y -) derivative elements correlated to $\eta_x(i, j)$. All the covariances between $\eta_x(i, j)$ and other points are represented.

$$\begin{aligned} \text{cov}(\eta_x(i, j), \eta_x(i, j+2)) &= \text{cov}(\eta_x(i, j), \eta_x(i, j-2)) \\ &= \text{cov}(\eta_x(i, j), \eta_y(i-1, j-1)) \\ &= \text{cov}(\eta_x(i, j), \eta_y(i+1, j+1)) \\ &= -\frac{\sigma^2}{4} \end{aligned}$$

which leads to a non-invertible \mathbf{R} matrix written as:

$$\mathbf{R} = \begin{pmatrix} C_{\eta_x \eta_x} & C_{\eta_x \eta_y} \\ C_{\eta_y \eta_x} & C_{\eta_y \eta_y} \end{pmatrix}$$

where $C_{\eta_x \eta_x}$ is the error covariance matrix between the x -derivative elements. Simplifying this covariance matrix by neglecting the correlation between x - and y - derivative elements leads to consider that $C_{\eta_x \eta_y} = 0$. This simplification is illustrated on Fig. 10. One consequence is that one element η_x (resp. η_y) is only correlated to elements of the same row (resp. column).

This approximation leads to consider the following block matrix $\tilde{\mathbf{R}}$:

$$\tilde{\mathbf{R}} = \begin{pmatrix} C_{\eta_x^{i=1} \eta_x^{i=1}} & 0 & \dots & 0 \\ 0 & C_{\eta_x^{i=2} \eta_x^{i=2}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & C_{\eta_x^{i=N} \eta_x^{i=N}} \end{pmatrix}$$

where $C_{\eta_x^{i=1} \eta_x^{i=1}}$ is the block corresponding to the covariance between x - derivative elements of the first row.

A2 Inversion of $\tilde{\mathbf{R}}$

Albeit the boundary elements, a block $C_{\eta_x^{i=k} \eta_x^{i=k}}$, of the matrix $\tilde{\mathbf{R}}$ can be separated in two blocks depending on j parity:

$$C_{\eta_x^{i=k} \eta_x^{i=k}} = \begin{pmatrix} C_{\eta_x^{j=odd} \eta_x^{j=odd}} & 0 \\ 0 & C_{\eta_x^{j=even} \eta_x^{j=even}} \end{pmatrix}$$

Taking into account that all blocks are identical we only need to invert four tridiagonal matrix $C_{\eta_x^{odd} \eta_x^{odd}}$, $C_{\eta_x^{even} \eta_x^{even}}$, $C_{\eta_y^{odd} \eta_y^{odd}}$, $C_{\eta_y^{even} \eta_y^{even}}$.

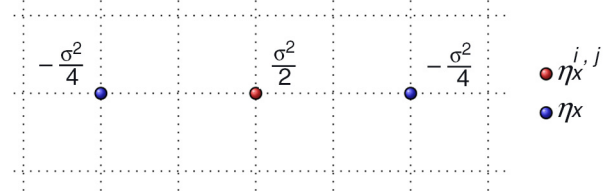


Fig. 10. Illustration of the simplification of the error covariance matrix. Only correlations between derivatives in the same direction are considered.

If $L=H=2N$, only one block matrix A (which size is $(N-1) \times (N-1)$) needs to be inverted to fill-in the full covariance matrix $\tilde{\mathbf{R}}$. This matrix reads:

$$\mathbf{M} = \sigma^2 \begin{pmatrix} 0.5 & -0.25 & 0 & \cdots & 0 \\ -0.25 & 0.5 & -0.25 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & 0 & -0.25 & 0.5 & -0.25 \\ \vdots & \vdots & 0 & -0.25 & 0.5 \end{pmatrix} \quad (\text{A3})$$

Its inverse exists and reads as:

$$\mathbf{M}^{-1} = \frac{4}{N\sigma^2} \mathbf{D}$$

with

$$\mathbf{D} = \begin{pmatrix} N-1 & N-2 & N-3 & \cdots \\ N-2 & 2*(N-2) & 2*(N-3) & \cdots \\ N-3 & 2*(N-3) & 3*(N-3) & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ 1 & 2 & \cdots & N-1 \end{pmatrix}$$

References

- Aouf, L., Lefèvre, J.-M. and Hauser, D. 2006. Assimilation of directional wave spectra in the wave model WAM: an impact study from synthetic observations in preparation for the SWIMSAT satellite mission. *J. Atmos. Ocean. Technol.* **23**, 448–463.
- Ba, S. and Fablet, R. 2010. Variational fronts tracking in sea surface temperature images. In: *17th IEEE International Conference on Image Processing (ICIP)*. Hong Kong, pp. 4117–4120.
- Barron, J. L., Fleet, D. J. and Beauchemin, S. S. 1994. Performance of optical flow techniques. *Int. J. Comput. Vision.* **12**(1), 43–77.
- Bormann, N., Saarinen, S. and Kelly, G. 2003. The spatial structure of observation errors in atmospheric motion vectors from geostationary satellite data. *Mon. Weather Rev.* **131**(4), 706–718.
- Candes, E., Demanet, L., Donoho, D. and Ying, L. 2006. Fast discrete curvelet transforms. *Multiscale Model. Simul.* **5**(3), 861–899.
- Cardinali, C. 2009. Monitoring the observation impact on the short-range forecast. *Q. J. Roy. Meteorol. Soc.* **135**(638), 239–250. DOI: 10.1002/qj.366.
- Chabot, V. 2014. *Étude de représentations parcimonieuses des statistiques d'erreur d'observation pour différentes métriques. Application à l'assimilation de données images* [Study of sparse representations of observation error statistics for different metrics. Application to image data assimilation]. PhD Thesis, Université de Grenoble.
- Courtier, P., Andersson, E., Heckley, W., Vasiljevic, D., Hamrud, M. and co-authors. 1998. The ECMWF implementation of three-dimensional variational assimilation (3d-Var). I: Formulation. *Q. J. Roy. Meteorol. Soc.* **124**(550), 1783–1807.
- Deckmyn, A. and Berre, L. 2005. A wavelet approach to representing background error covariances in a limited-area model. *Mon. Weather Rev.* **133**(5), 1279–1294.
- Desroziers, G., Berre, L., Chabot, V. and Chapnik, B. 2009. A posteriori diagnostics in an ensemble of perturbed analyses. *Mon. Weather Rev.* **137**, 3420–3436.
- Desroziers, G., Berre, L., Chapnik, B. and Poli, P. 2005. Diagnosis of observation, background and analysis-error statistics in observation space. *Q. J. Roy. Meteorol. Soc.* **131**(613), 3385–3396.
- D'Ovidio, F., Taillandier, V., Mortier, L. and Taupier-Letage, I. 2009. Lagrangian validation of the Mediterranean mean dynamic topography by extraction of tracer frontal structures. *Mercator Ocean Q. Newsl.* **32**, 24–32.
- Fadili, J., Starck, J.-L., Elad, M. and Donoho, D. 2010. MCALab: reproducible research in signal and image decomposition and inpainting. *Comput. Sci. Eng.* **12**(1), 44–63.
- Fisher, M. 2003. Background error covariance modelling. *Seminar on Recent Development in Data Assimilation for Atmosphere and Ocean*. ECMWF, Reading, UK, pp. 45–63.
- Flór, J.-B. and Eames, I. 2002. Dynamics of monopolar vortices on a topographic beta-plane. *J. Fluid Mech.* **456**(1), 353–376.
- Gaultier, L., Verron, J., Brankart, J.-M., Titaud, O. and Brasseur, P. 2013. On the inversion of submesoscale tracer fields to estimate the surface ocean circulation. *J. Mar. Syst.* **126**, 33–42.
- Gilbert, J. C. and Lemaréchal, C. 1989. Some numerical experiments with variable-storage quasi-newton algorithms. *Math. Program.* **45**(1–3), 407–435.
- Horn, B. and Schunck, B. 1981. Determining optical flow. *Artif. Intell.* **17**, 185–203.
- Köpken, C., Kelly, G. and Thépaut, J.-N. 2004. Assimilation of Meteosat radiance data within the 4d-var system at ECMWF: assimilation experiments and forecast impact. *Q. J. Roy. Meteorol. Soc.* **130**(601), 2277–2292.
- Korotaev, G., Huot, E., Le Dimet, F.-X., Herlin, I., Stanichny, S. and co-authors. 2008. Retrieving ocean surface current by 4-d variational assimilation of sea surface temperature images. *Remote Sens. Environ.* **112**(4), 1464–1475.
- Le Dimet, F.-X. and Talagrand, O. 1986. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A.* **38**, 97–110.
- Liu, Z.-Q. and Rabier, F. 2002. The interaction between model resolution, observation resolution and observation density in data assimilation: a one-dimensional study. *Q. J. Roy. Meteorol. Soc.* **128**(582), 1367–1386.
- Lucas, B. and Kanade, T. 1981. An iterative image registration technique with an application to stereovision. *IJCAI'81 Proceedings of the 7th International Joint Conference on Artificial Intelligence*. Vancouver, Canada, pp. 674–679.
- Matricardi, M., Chevallier, F., Kelly, G. and Thépaut, J.-N. 2004. An improved general fast radiative transfer model for the assimilation of radiance observations. *Q. J. Roy. Meteorol. Soc.* **130**(596), 153–173.
- Michel, Y. 2011. Displacing potential vorticity structures by the assimilation of pseudo-observations. *Mon. Weather Rev.* **139**(2), 549–565.
- Nguyen van Yen, R., Farge, M. and Schneider, K. 2012. Scale-wise coherent vorticity extraction for conditional statistical

- modeling of homogeneous isotropic two-dimensional turbulence. *Physica D*. **241**, 186–201.
- Nieman, S., Menzel, P., Hayden, C., Gray, D., Wanzong, S. and co-authors. 1997. Fully automated cloud-drift winds in NESDIS operations. *Bull. Am. Meteorol. Soc.* **78**, 1121–1133.
- Pannekoucke, O. 2009. Heterogeneous correlation modelling based on the wavelet diagonal assumption and on the diffusion operator. *Mon. Weather Rev.* **137**(9), 2995–3012.
- Papadakis, N. and Mémin, E. 2007. A variational framework for spatiotemporal smoothing of fluid motions. *Lect. Notes Comput. Sci.* **4485**, 603–615.
- Papadakis, N. and Mémin, E. 2008. Variational assimilation of fluid motion from image sequence. *SIAM J. Imaging. Sci.* **1**, 343–363.
- Schmetz, J., Holmlund, K., Hoffman, J., Strauss, B., Mason, B. and co-authors, 1993. Operational cloud-motion winds from Meteosat infrared images. *J. Appl. Meteorol.* **32**(7), 1206–1225.
- Souopgui, I. 2010. *Assimilation d'images pour les fluides géophysiques* [Image assimilation for geophysical fluids]. PhD Thesis, Université de Grenoble.
- Stewart, L. M., Dance, S. and Nichols, N. K. 2013. Data assimilation with correlated observation errors: experiments with a 1-d shallow water model. *Tellus A*. **65**, 19546. DOI: 10.3402/tellusa.v65i0.19546.
- Titau, O., Vidard, A., Souopgui, I. and Le Dimet, F.-X. 2010. Assimilation of image sequences in numerical models. *Tellus A*. **62**(1), 30–47.
- Titau, O., Brankart, J.-M. and Verron, J. 2011. On the use of finite-time Lyapunov exponents and vectors for direct assimilation of tracer images into ocean models. *Tellus A*. **63**(5), 1038–1051.
- Vannucci, M. and Corradi, F. 1999. Covariance structure of wavelet coefficients: theory and models in a Bayesian perspective. *J. R. Statist. Soc. B*. **61**(4), 971–986.
- Weaver, A. and Courtier, P. 2001. Correlation modelling on the sphere using a generalized diffusion equation. *Q. J. Roy. Meteorol. Soc.* **127**, 1815–1846.